Natural Language Processing: Part II Overview of Natural Language Processing (L90): ACS

Ann Copestake

Computer Laboratory
University of Cambridge

October 2014

Outline of today's lecture

Lecture 1: Introduction

- Overview of the course
- Why NLP is hard
- Scope of NLP
- A sample application: sentiment classification
- More NLP applications
- NLP subtasks

Part II / ACS

- Part II
  - 12 lectures, assessed by exam questions (as in previous years).
  - Supervisions.
- ACS
  - L90: new module for this year.
  - Overview of NLP: other modules go into much greater depth: L90 intended for people with no substantial background in NLP.
  - Same 12 lectures as Part II, plus extended practical organised by Simone Teufel (with demonstrators).
  - No supervisions, Q&A session(s) will be offered.

Also note:

- Lecture notes.
- No notes for lectures 11 (guest lecture) and 12: not directly examinable.
- Slides: on web page in advance, but possible (slight) differences to slides used in lecture.
- Exercises: pre-lecture and post-lecture.
- Glossary in lecture notes.
- Webpage with links to demos etc.
- Recommended Book: Jurafsky and Martin (2008).
- Linguistics background: Bender (2013).

NLP and linguistics

NLP: the computational modelling of human language.

1. Morphology — the structure of words: lecture 2.
2. Syntax — the way words are used to form phrases: lectures 3, 4 and 5.
3. Semantics
   - Compositional semantics — the construction of meaning based on syntax: lecture 6.
   - Lexical semantics — the meaning of individual words: lecture 7 and 8.
4. Pragmatics — meaning in context: lecture 9.
5. Language generation — lecture 10.
6. Humans vs machines — lecture 11.

Querying a knowledge base

User query:

- Has my order number 4291 been shipped yet?

Database:

ORDER

Order number	Date ordered	Date shipped
4290	2/2/13	2/2/13
4291	2/2/13	2/2/13
4292	2/2/13	

USER: Has my order number 4291 been shipped yet?
DB QUERY: order(number=4291,date_shipped=?)
RESPONSE: Order number 4291 was shipped on 2/2/13

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

- Do you sell Sony laptops and disk drives?
- Do you sell (Sony (laptops and disk drives))?
- Do you sell (Sony laptops) and disk drives)?

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

- Do you sell Sony laptops and disk drives?
- Do you sell (Sony (laptops and disk drives))?
- Do you sell (Sony laptops) and disk drives)?

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

- Do you sell Sony laptops and disk drives?
- Do you sell (Sony (laptops and disk drives))?
- Do you sell (Sony laptops) and disk drives)?

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

- Do you sell Sony laptops and disk drives?
- Do you sell (Sony (laptops and disk drives))?
- Do you sell (Sony laptops) and disk drives)?

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

- Do you sell Sony laptops and disk drives?
- Do you sell (Sony (laptops and disk drives))?
- Do you sell (Sony laptops) and disk drives)?

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

- Do you sell Sony laptops and disk drives?
- Do you sell (Sony (laptops and disk drives))?
- Do you sell (Sony laptops) and disk drives)?

Wouldn't it be better if . . . ?

The properties which make natural language difficult to process are essential to human communication:

- Flexible
- Learnable but compact
- Emergent, evolving systems

Synonymy and ambiguity go along with these properties.

Natural language communication can be indefinitely precise:

- Ambiguity is mostly local (for humans)
- Semi-formal additions and conventions for different genres

Wouldn't it be better if . . . ?

The properties which make natural language difficult to process are essential to human communication:

- Flexible
- Learnable but compact
- Emergent, evolving systems

Synonymy and ambiguity go along with these properties.
Natural language communication can be indefinitely precise:

- Ambiguity is mostly local (for humans)
- Semi-formal additions and conventions for different genres

Some NLP applications

- spelling and grammar checking
- predictive text
- optical character recognition (OCR)
- screen readers
- augmentative and alternative communication
- machine aided translation
- lexicographers' tools
- information retrieval
- document classification
- document clustering
- information extraction
- sentiment classification
- text mining

More NLP applications . . .

- question answering
- summarization
- text segmentation
- exam marking
- language teaching
- report generation
- machine translation
- natural language interfaces to databases
- email understanding
- dialogue systems

Opinion mining: what do they think about me?

- Task: scan documents (webpages, tweets etc) for positive and negative opinions on people, products etc.
- Find all references to entity in some document collection: list as positive, negative (possibly with strength) or neutral.
- Construct summary report plus examples (text snippets).
- Fine-grained classification:
  e.g., for phone, opinions about: overall design, display, camera.

LG G3 review (Guardian 27/8/2014)

The shiny, brushed effect makes the G3's plastic design looks deceptively like metal. It feels solid in the hand and the build quality is great — there's minimal give or flex in the body. It weighs 149g, which is lighter than the 160g HTC One M8, but heavier than the 145g Galaxy S5 and the significantly smaller 112g iPhone 5S.

The G3's claim to fame is its 5.5in quad HD display, which at 2560x1440 resolution has a pixel density of 534 pixels per inch, far exceeding the 432ppi of the Galaxy S5 and similar rivals. The screen is vibrant and crisp with wide viewing angles, but the extra pixel density is not noticeable in general use compared to, say, a Galaxy S5.

Sentiment classification: the research task

- Full task: information retrieval, cleaning up text structure, named entity recognition, identification of relevant parts of text. Evaluation by humans.
- Research task: preclassified documents, topic known, opinion in text along with some straightforwardly extractable score.
- Movie review corpus (Pang et al 2002): strongly positive or negative reviews from IMDb, 50:50 split, with rating score.

IMDb: An American Werewolf in London (1981)

Rating: 9/10

> *Ooooo. Scary.*
> *The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.*
> *In Summary: Top banana*

Bag of words technique

- Treat the reviews as collections of individual words.
- Classify reviews according to positive or negative words.
- Could use word lists prepared by humans, but machine learning based on a portion of the corpus (training set) is preferable.
- Use human rankings for training and evaluation.
- Pang et al, 2002: Chance success is 50% (corpus artificially balanced), bag-of-words gives 80%.

Some sources of errors for bag-of-words

- Negation:

  Ridley Scott has never directed a bad film.

- Overfitting the training data:
  e.g., if training set includes a lot of films from before 2005, *Ridley* may be a strong positive indicator, but then we test on reviews for ‘Kingdom of Heaven’?
- Comparisons and contrasts.

Contrasts in the discourse

This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.

More contrasts

AN AMERICAN WEREWOLF IN PARIS is a failed attempt . . . Julie Delpy is far too good for this movie. She imbues Serafine with spirit, spunk, and humanity. This isn't necessarily a good thing, since it prevents us from relaxing and enjoying AN AMERICAN WEREWOLF IN PARIS as a completely mindless, campy entertainment experience. Delpy's injection of class into an otherwise classless production raises the specter of what this film could have been with a better script and a better cast . . . She was radiant, charismatic, and effective . . .

Doing sentiment classification ‘properly’?

- Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with what, tense . . .
- Lexical semantics:
  are words positive or negative in this context? Word senses (e.g., *spirit*)?
- Pragmatics and discourse structure:
  what is the topic of this section of text? Pronouns and definite references.
- Getting all this to work well on arbitrary text is very hard.
- Ultimately the problem is AI-complete, but can we do well enough for NLP to be useful?

IR, IE and QA

- Information retrieval: return documents in response to a user query (Internet Search is a special case)
- Information extraction: discover specific information from a set of documents (e.g. company joint ventures)
- Question answering: answer a specific user question by returning a section of a document:

  What is the capital of France?
  Paris has been the French capital for many centuries.

MT

- Earliest attempted NLP application.
- High quality only if the domain is restricted (or with very close languages: e.g., Swedish-Danish).
- Utility greatly increased in 1990s with increase in availability of electronic text.
- Good applications for bad MT . . .
- Spoken language translation is viable for limited domains.

Natural language interfaces and dialogue systems

All rely on a limited domain:

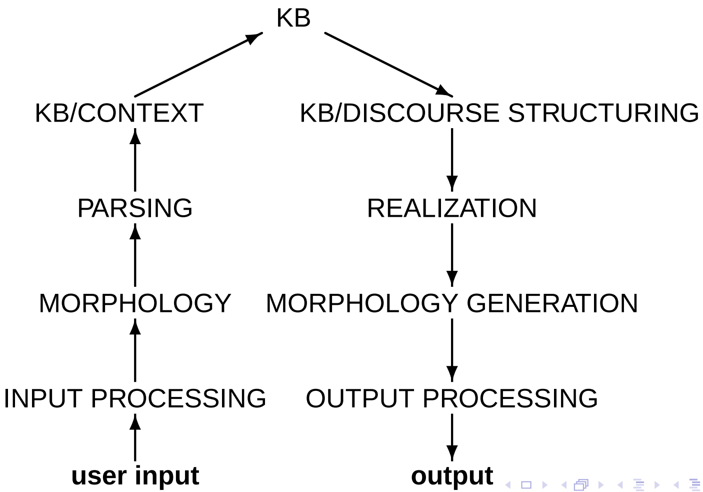
- LUNAR: classic example of a natural language interface to a database (NLID): 1970–1975
- SHRDLU: (text-based) dialogue system: 1973
- Current spoken dialogue systems

Limited domain allows disambiguation: e.g., in LUNAR, *rock* had one sense.

NLP subtasks

- input preprocessing: speech recognizer, text preprocessor or gesture recognizer.
- morphological analysis (2)
- part of speech tagging (3)
- parsing: this includes syntax and compositional semantics (4, 5, 6)
- disambiguation, inference (6, 7, 8)
- context processing (9)
- discourse structuring (10)
- realization (10)
- morphological generation (2)
- output processing: text-to-speech, text formatter, etc.

Subtasks in natural language interface to a knowledge base



General comments

- Even ‘simple’ applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $<$ 100% perfect can be useful.
- Aids to humans are easier than replacements for humans.
- NLP interfaces compete with non-language approaches.
- Shallow processing on arbitrary input or deep processing on narrow domains.
- Limited domain systems require extensive and expensive expertise to port or large amounts of data (also expensive).
- External influences on NLP are very important.